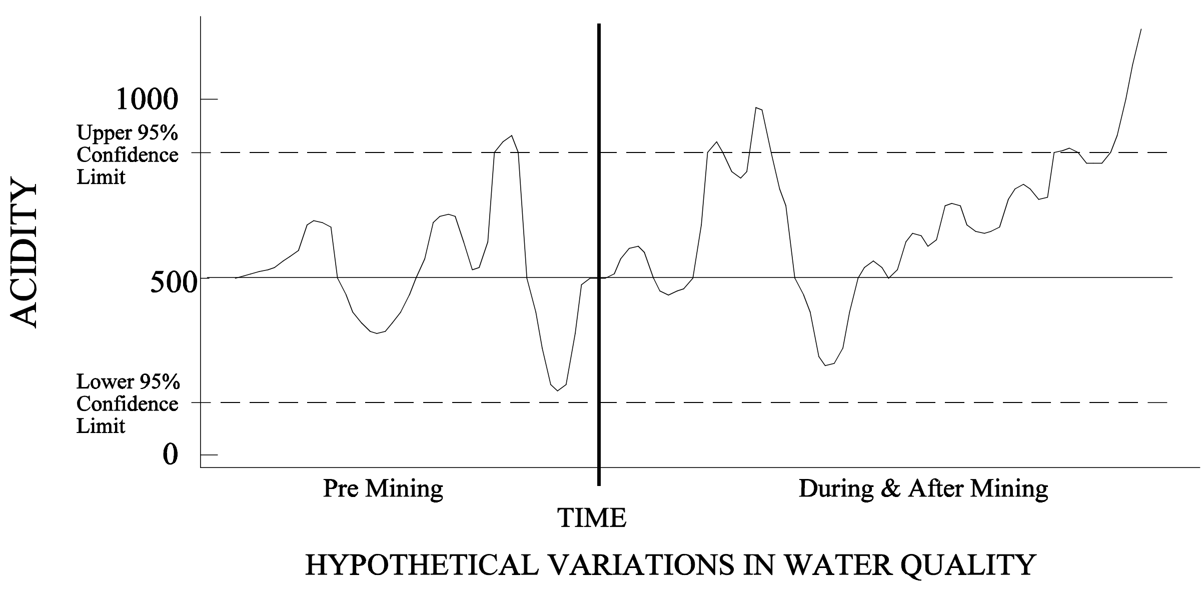
Chapter 2: Statistical Analysis of Mine Drainage Data

If discharges from abandoned mines did not vary in flow or water quality parameters through time, it would not be necessary to use statistics to determine the baseline pollution loads of a remining site. In fact, the baseline determination would involve little effort, in terms of representative sampling and chemical and data analyses. A mine operator or regulatory agency could simply collect one sample to initially establish the baseline flow and water quality, and then collect a second sample at some later time before remining commences to document that the flow and water quality parameters do not vary through time. However, abandoned mine discharges typically vary significantly in flow and/or quality throughout the water year, and it is necessary to use statistics to quantify and explain these variations. Data representing this variation and the statistical analysis of such variation are presented in the succeeding chapters of this report. This chapter provides an introduction to the statistical methods that may be employed in determining the baseline pollution load.

Variation

The fundamental problem to be addressed in determining baseline pollution load is how to statistically summarize the natural variations in flow and water quality parameters before remining commences, in order to enable the separation of mining-induced changes in pollution load from natural seasonal variations in pollution load during and following remining operations. This problem is depicted in Figure 2.1, which shows hypothetical variations in acid load of an abandoned mine discharge before (pre), during, and after remining. It is important to note that Figure 2.1 is presented for graphical description of statistical triggers only and that the after-mining scenario represented in this figure is atypical. In almost all cases, remining will improve water quality. Whatever the case, water quality data should be plotted and statistically analyzed to determine whether adverse effects have occurred.

Figure 2.1: Example of Acid Load Variation Before, During, and After Remining



In Figure 2.1, the acid load varied greatly before remining commenced ranging from nearly zero pounds of acidity per day (lbs/day) to nearly 1,000 lbs/day. Observe that before remining, the discharge usually varies somewhat symmetrically above and below the central value of 500 lbs/day (central tendency) and that the variations are generally contained between the values of 50 lbs/day and 950 lbs/day which have been labeled the lower and upper control levels. Also note that the acid load was higher than the upper control level on one or two occasions before remining commenced. However, during remining the acid load was above the upper control level much more frequently, while the acid load is still varying somewhat symmetrically above and below the central tendency value for at least the first two years during remining. Finally, during the last three years of remining and following the completion of remining, the acid load still varies above and below the central tendency value, but there appears to be a trend of increasing acidity between the central tendency value and the upper control level.

In order to determine baseline pollution load, it is necessary to statistically analyze the data to find a measure of central tendency (e.g., mean or median) and a measure of the patterns of variation or the dispersion of the individual observations (i.e., samples around the central tendency as shown in Figure 2.1). In order to separate mining-induced changes in pollution load from natural seasonal variations, it is necessary to develop a statistical mechanism to determine when variations in the pollution load are out of control; that is, when significant deviations from the pre-remining baseline have occurred which can be attributed to factors other than natural seasonal variations (e.g., problems within remining operations, unrepresentative baseline, inappropriate monitoring).

There are two types of variation in pollution load which are of interest in evaluating monitoring data during and after remining in order to determine whether the variations are out of control from the established baseline conditions.

- <u>Dramatic Trigger</u> - The first and most obvious pattern of variation occurs when there are a series of extreme events which consistently exceed the upper control level as shown in Figure 2.1 during the first two years of remining. During this time, the variation pattern indicates a sudden and dramatic increase in pollution load which may be attributed to remining, and which is referred to as the dramatic trigger.

- <u>Subtle Trigger</u> - The second pattern of variation of concern is a trend of gradually increasing pollution load (as shown in the right side of Figure 2.1,) where the general pattern of acid load observations is increasing above the baseline central tendency value for several years without ever exceeding the upper control level. In this case, when the central tendency values are calculated for each water year during remining, a corresponding gradual increase in central tendency values will be detected until a significant difference exists between the baseline central tendency and a central tendency calculated for a water year after remining has commenced. As this second pattern of variation is much less dramatic than the first, and takes much more time and effort to detect, it is referred to as the subtle trigger.

The reason that these two patterns of variation are referred to as triggers is that they can be used to set off or initiate the requirement for a mine operator to treat a pre-existing discharge to a numerical effluent limitation. In issuing a remining permit, the regulatory authority makes a determination that the site can be mined without causing additional pollution, and that the pollution abatement plan in the permit application demonstrates that the existing baseline pollution load will be reduced. The mine operator and the regulatory authority anticipate environmental improvement through remining without the need to treat the pre-existing discharge. However, the possibility exists that degradation of the discharge may occur, temporarily or permanently, as the result of remining if the pollution abatement plan is not implemented as required or if unforeseen circumstances develop.

If fair and reasonable consideration is given to the concerns of the mine operator and protection of the environment, the treatment triggers must be carefully established so that they are: (a) not set off prematurely or erroneously, adversely affecting the mine operator, or (b) set off too late resulting in additional mine drainage pollution without treatment. Even the most thorough representative sampling program of a given water year may not capture the most extreme events, because the worst storm (flood) and the most severe drought are rare events and do not occur in every water year. Although it is unreasonable to require a mine operator to collect baseline water samples until the 100 year storm event or a significant drought are captured, it is also unreasonable to require the mine operator to commence treatment the first time that the extreme event or upper control level of the baseline is exceeded. The reagent costs alone for treating some pre-existing pollutional discharges can be several hundred dollars per day and the total cost of building a treatment plant can be more than one million dollars. Costs for treatment of some worst case post mining discharges in the State of Pennsylvania were as high as $ 700 /day (hydrated lime). Cost for construction of these discharges were greater than $ 2.1 million.

Conversely, the regulatory authority is not fulfilling its environmental protection mandate if the upper control level and extreme events of baseline are routinely being exceeded and the additional mine drainage pollution effects are obvious, but treatment has not yet been required because statistical analysis of the water year has not been completed. In light of these concerns, problems that need to be resolved statistically with respect to the dramatic and subtle triggers are:

- Dramatic trigger - how high should the upper control level or tolerance level be, and how many excursions above this upper level are tolerable before it is determined that the system is out of control and treatment of the discharge must be initiated.

- Subtle trigger - how much deviation from the baseline central tendency value is tolerable in succeeding water years before it can be determined that a significant difference exists.

Both of these problems may be addressed statistically with a relatively simple quality control approach to the data.

<u>Normal Distribution</u>
The quality control approach used in this report and much of statistical work in general, is dependent upon the frequency distribution of the sample data. It is important to collect representative samples, because it is usually impossible or impractical to measure and analyze

the entire population of the parameter being studied. Whether the samples represent variation in a single point through time (e.g., seasonal variations in the acidity of an abandoned mine discharge) or spatial variations in a parameter of interest (e.g., variations in the mean acidity of surface mine discharges from the lower Kittanning coal seam of 200 sites in western Pennsylvania), one of the first steps of statistical analysis, typically, is to plot the frequency distribution of the data. According to Sir Ronald A. Fisher (1970), the founder of many important statistical advances since the 1920's:

> "The idea of an infinite **population** distributed in a **frequency distribution** in respect of one or more characters is fundamental to all statistical work. From a limited experience, for example, of individuals of a species, or of the weather of a locality, we may obtain some idea of the infinite hypothetical population from which a sample is drawn, and so of the probable nature of future samples to which our conclusions are to be applied. If a second sample belies this expectation we infer that it is, in the language of statistics, drawn from a different population; that the treatment to which the second sample of organisms had been exposed did in fact make a material difference, or that the climate (or the methods of measuring it) had materially altered. Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first." (p. 41)

Fisher (1970) also states:

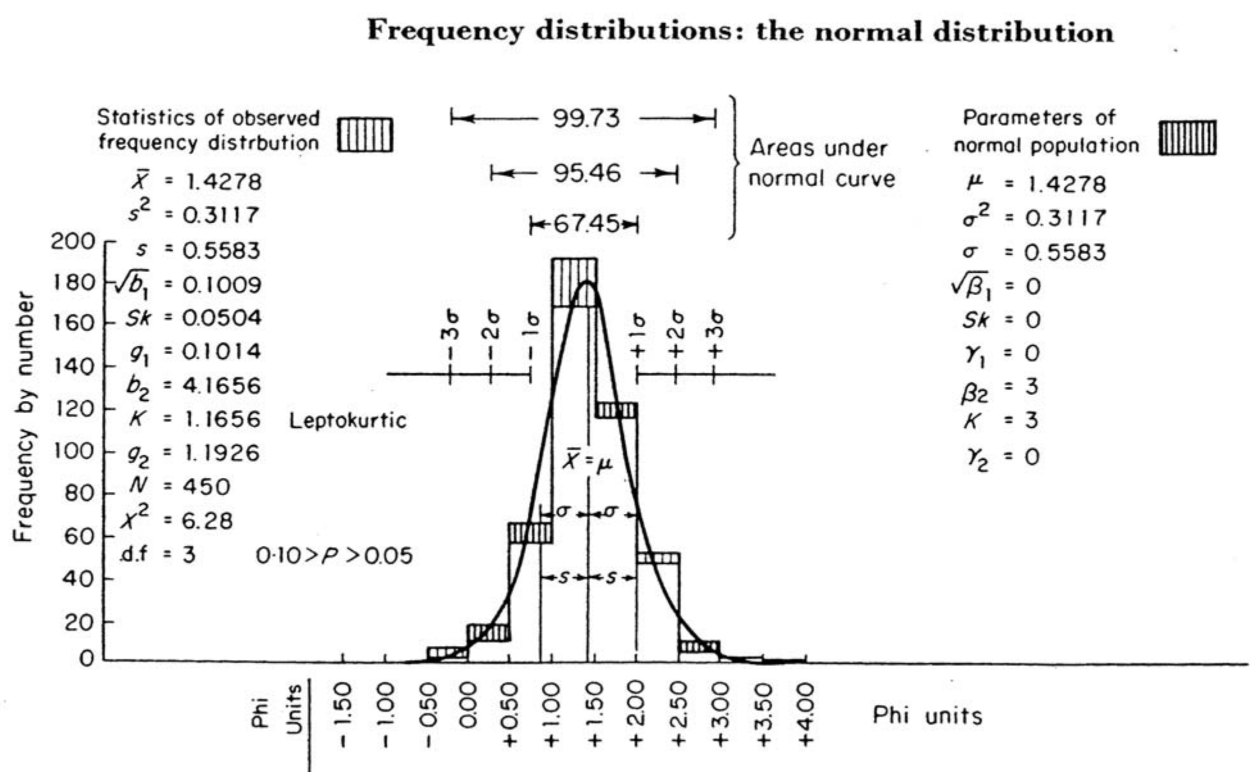
> "Statistics may be regarded as (i) the study of **populations**, (ii) the study of **variation**, (iii) the study of methods of the **reduction of data** (p. 1)… [and] ….A **statistic** is a value calculated from an observed sample with a view to characterizing the population from which it is drawn." (p. 41)

The frequency distribution is a graphical summary of the sample data, and its shape and accompanying summary statistics enable a greater understanding of how a variate behaves. This understanding is gained through comparison of the frequency distribution of observed data to the shape and characteristics of a known mathematical or theoretical distribution, such as the normal distribution or binomial distribution. The normal distribution shown in Figure 2.2 is the most widely known and most useful frequency distribution. It is also known as the Gaussian error curve or bell-shaped curve.

The key statistical parameters of a normal frequency distribution are the mean, as the measure of central tendency (i.e., shown as $\overline{X}$ in Figure 2.2), and the standard deviation or the variance, as the measure of variation or dispersion (i.e., the standard deviation is shown as $\hat{\sigma}$ in Figure 2.2). The mean is the arithmetic average of the data, which is computed by dividing the sum of all of the observations by the total number of observations. The variance is the sum of the squares of the deviations of all of the observations from the mean. The standard deviation is the positive square root of the variance.

According to Fisher (1970) and Griffiths (1967), the sample mean ($\overline{X}$) and standard deviation ($\hat{\sigma}$) determined from a random sample are best estimators of the corresponding population mean (μ) and standard deviation (σ) in a normal distribution. They are defined as "best estimators" because these statistics are consistent, efficient, and sufficient, and in the most desirable outcome, unbiased as well. A primary goal of parametric statistical analysis is that the statistical estimators (e.g., mean and standard deviation) of the sample distribution will converge on the population parameters (i.e., the true mean and variance of the entire population). Most parametric statistical methods, including tests of significance, are based upon: (a) the use of the mean and the standard deviation as best sufficient statistical estimators, and (b) the assumption that the sample data are normally distributed.

Figure 2.2: Example of Normal Distribution



In addition, probability statements, which are used in significance testing, quality control techniques and other statistical methods, are frequently based upon some special properties of the normal distribution (see Griffiths, 1967, pp. 263 – 267). The area under the curve of the normal distribution in the interval between the mean minus one standard deviation and the mean plus one standard deviation (as shown in Figure 2.2, from Griffiths, 1967, p. 259) is 67.45%, while 95.46% of the area of the normal distribution is contained in the interval of the mean plus and minus two standard deviations (i.e., $\overline{X}$ +/- 2 $\hat{\sigma}$). Therefore, from the table of areas of the normal distribution it may be stated that 95% of the area of the distribution will be contained in the interval of $\overline{X}$ +/- 1.96$\hat{\sigma}$, Griffiths, 1967, p. 265).

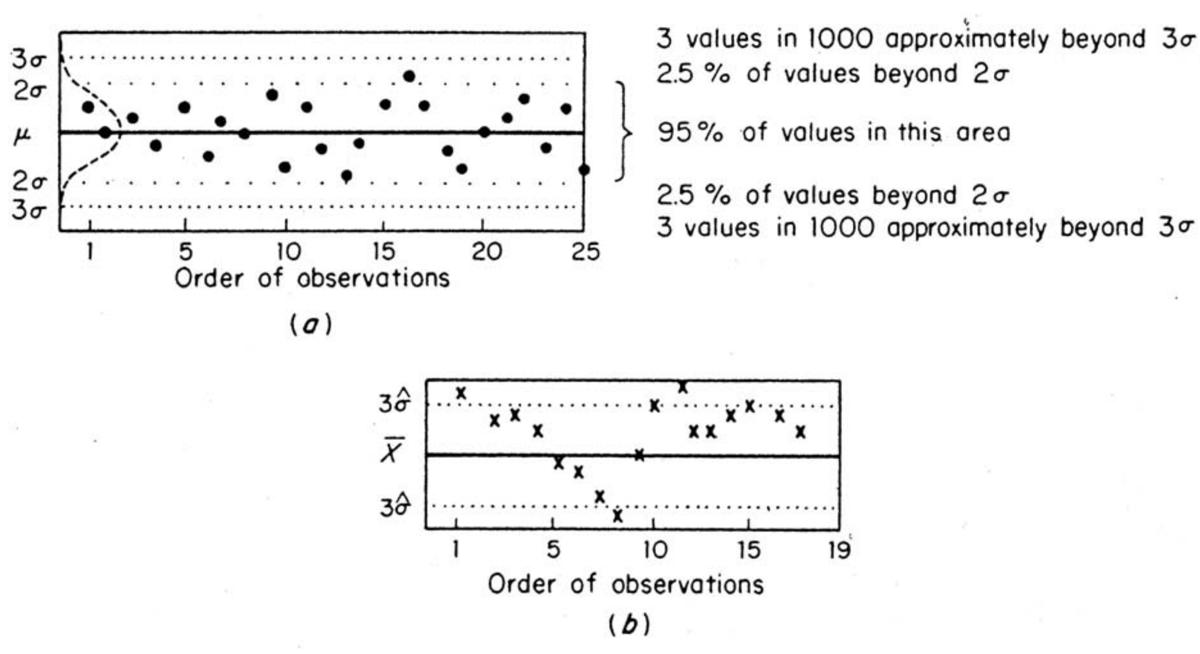
<u>Quality Control - Normal Distribution</u>

The type of statistical analysis known as quality control was largely developed by Shewhart (1931, 1939) and others to evaluate tolerable amounts of variation in manufacturing processes. Since then, the quality control approach has been applied to many other fields of study. Many of the variates studied in very large samples, such as the number of defective light bulbs produced by a manufacturing process, were empirically shown to closely approximate a normal distribution. Consequently, the most typical applications of quality control statistics involve a normal distribution.

The frequency distribution of the data is essentially arranged along the vertical axis of the quality control graph as shown in Figure 2.3. The actual histogram of value classes is typically omitted from the graph. The mean of the data set, or grand mean of the means of sets of observations, is usually plotted as the measure of central tendency. Quality control levels, known as confidence intervals, are established at plus and minus two or three standard deviations from the mean. Individual observations through time, or comparisons of sets of data representing variations in operator performance, are then plotted along the horizontal axis in order to evaluate the patterns of variation in these observations with respect to the confidence intervals around the mean. As 95.46% of the area of the normal frequency distribution is contained in the interval of the mean +/- two standard deviations, it is expected that approximately 95 out of 100 observations will occur within the confidence intervals.

Figure 2.3: Example of a Quality Control Graph (Griffiths, 1967, p. 318)

According to Griffiths (1967):



> If the observations are in control, they will fluctuate randomly around the mean value, and some 5 in 100 will fall outside the 2σ limits or 3 in 1,000 will be expected outside the 3σ limits. If a number of observations exceeding these expectations fall outside the control limits, these observations may be looked upon as not belonging to the same population.
>
> The fact that the observations are ordered, however, permits an extension of this generalization; the observations need not fall outside the limits to indicate lack of control. For example, in an industrial process which produces a homogeneous product when the process is in control, the wear on a machine may develop gradually; then a series of observations will show a systematic trend, i.e., the characteristic of random variation will be lost. Such trends arise from a systematic bias, and it is customary to search for an "assignable cause" and remove the interfering source of variation by replacing worn parts in the machine and so on. The systematic trend may appear and the presence of such a trend warns the observer that his observations are not in control (p. 318).

<u>Asymmetric Distribution</u>

A major problem that is frequently encountered in the statistical analysis of water quality parameters and many other variates of natural systems behavior is that the sample data are not normally distributed (Reimann and Filzmoser, 2000). In analyzing the concentrations of acidity, iron, sulfates, or other water quality parameters of abandoned mine discharges, ground water and surface water, it is typical to have many small valued observations in the data set and a few very large values representing extreme events. This type of behavior appears to be relatively common for variates which are related to seasonal variations, climatic effects, and geological or biological systems (for examples, see Aitcheson & Brown (1973), Griffiths (1967), and Krumbein & Graybill (1965)). The frequency distribution for this type of variate is highly asymmetric as shown in Figure 2.4. The few very high valued observations cause the frequency distribution to have a long tail toward the high extremes, which is a condition of asymmetry termed positive skewness.

Figure 2.4 : Stem-and-leaf of Discharge (Example of Asymmetric Distribution)

```
N = 81
Leaf Unit = 0.10

40      0     0111100000222222222222222223333344444444
(24)    0     555555555566666777888899
17      1     0123334
10      1     69

8       2     13
6       2     69
4       3     014
1       3
1       4
1       4
1       5     0
```

In the normal frequency distribution, the values are symmetrically distributed around the mean, and the mean and standard deviation are best statistical estimators of the population. In a highly skewed frequency distribution, the mean may not be the best estimator of central tendency, and the standard deviation may not be the best measure of dispersion.

For example, the few extreme values bias the mean toward the high values, and 95% of the area of the curve is not contained within ± 2 standard deviations from the mean. In cases where the frequency distribution is not normal, the concept of the quality control approach may still be pursued, but data analysis adjustments must be made to either: (a) transform the observed frequency distribution to approximate normality, or (b) employ different statistics (e.g., use of the median instead of the mean) in the quality control technique.

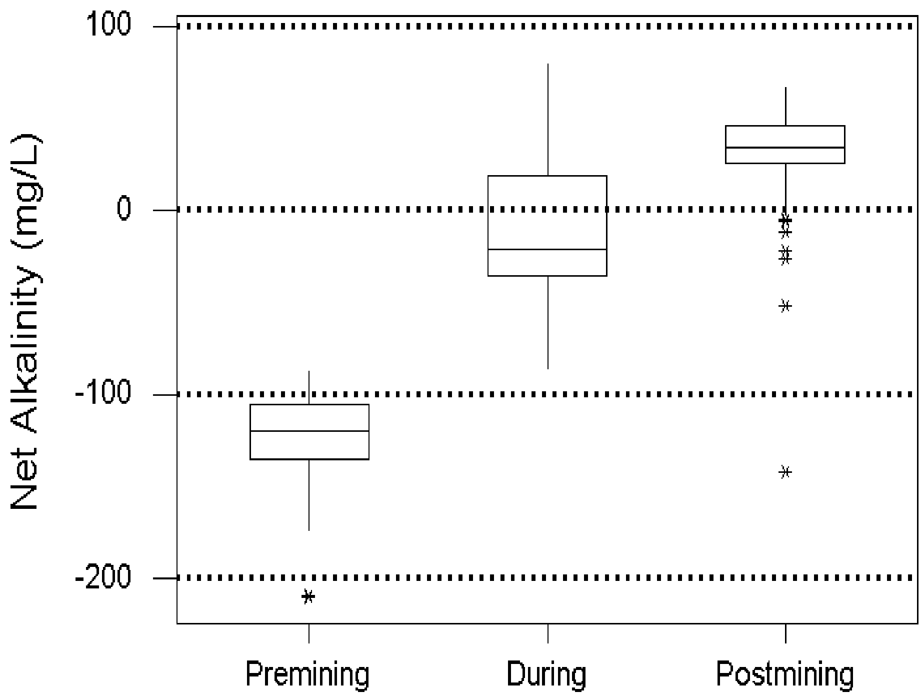
The logarithmic transformation of the data is usually the most effective transformation to reduce positive skewness in the frequency distribution. The lognormal distribution had been extensively described by Aitcheson and Brown (1973) and examples of lognormal behavior of variates are found in Griffiths (1967), Krumbein and Graybill (1965) and other sources. However, a logarithmic transformation of the raw data will not solve all problems of asymmetry or other conditions of non-normality of the frequency distribution. Additional information on transformations of data is described later in this chapter and in Box and Cox (1964), Griffiths (1967, p. 306), and Krumbein and Graybill (1965, p. 216).

In order to evaluate different statistics that may be applicable to the quality control approach, it is necessary to explain and differentiate nonparametric statistics, distribution free statistics, and order statistics. It is also necessary to compare exploratory data analysis with confirmatory data analysis.

- Conventional Parametric Statistical Analysis - statistical estimators, such as the mean and standard deviation, are used to approximate the corresponding parameters of the population, the true mean and variance.
- Nonparametric Statistical Analysis - tests of significance are performed without depending on the constraints of a known frequency distribution and the parameters of that known frequency distribution (e.g., the mean and variance of the normal distribution). Nonparametric statistical tests are also used where the scale level of the data are only nominal or ordinal, rather than on interval or ratio scales used in more rigorous statistical analyses. However, if the data conform to a known frequency distribution, there are parameters for that distribution.
- Distribution-free Statistics - is used to describe statistical analyses where parameters are estimated independently of the shape of the frequency distribution, such as the use of the chi-square statistic to test the class by class departure from the expected value.
- Order Statistics - is applied to statistical analyses where the shape of the frequency distribution is important, but is evaluated less rigorously than in conventional parametric statistical analyses. In order statistics, the median is typically used as the measure of central tendency instead of the mean, and quartiles or related values are typically used to measure dispersion, the spread of values about the median, or the shape of the distribution. The position of the median in an ordered set of observations is the middlemost position. For

example, when 15 values are ordered from low to high, the depth (position) of the median is at the (N+1)/2 position = 8th position. The position of the two quartiles (Q_1, Q_3) in this ordered set is halfway between the median and the extremes (e.g., lower quartile (Q_1) is at the (8+1)/2 = midway between the 4th and 5th observation). The quartile (Q_1 in this case), is found by counting in from either extreme to the 4th observation. The quartiles essentially divide the frequency distribution into fourths, so that half of the values in the distribution are contained in the interval between the lower quartile and the upper quartile as shown in Figure 2.5 (i.e., within box). Other values of spread or dispersion are similarly determined based upon their rank or order in the frequency distribution.

Figure 2.5: Net Alkalinity Boxplot for Fisher Mine Site Discharge (from U.S. EPA Coal



Remining Statistical Support Document, March 2000, EPA-821-B-00-001)

As a final note on the relationships of the various frequency distributions discussed herein and elsewhere (see Fisher (1970), Fisher (1973), Griffiths (1967), Krumbein and Graybill (1965) and Tukey (1977)), regardless of the shape of the frequency distributions in samples of water quality parameters or almost any other variable of interest, the distribution of the means of sample sets or means of repeated sampling efforts tend to be normally distributed. Generally, frequency distributions tend toward normality as the number of observations in the sample set becomes very large (i.e., greater than 1 million observations). However, most samples of mine drainage data used in remining permitting and monitoring will contain a relatively small number of observations (i.e., less than 30).

<u>Exploratory and Confirmatory Data Analyses</u>
Most of the statistical analyses discussed thus far, especially significance tests, can be included in the realm of confirmatory data analysis rather than exploratory data analysis.

According to Tukey (1977):

> The principles and procedures of what we call confirmatory data analysis are both widely used and one of the great intellectual products of our century. In their simplest form, these principles and procedures look at a sample -- and at what that sample has told us about the population from which it came -- and assess the precision with which our inference from sample to population is made. We can no longer get along without confirmatory data analysis. **But we need not start with it….**(p. vi)
>
> Once upon a time, statisticians only explored. Then they learned to confirm exactly – to confirm a few things exactly, each under very specific circumstances. As they emphasized exact confirmation, their techniques inevitably became less flexible. The connection of the most used techniques with past insights was weakened. Anything to which a confirmatory procedure was not explicitly attached was described as "mere descriptive statistics," no matter how much we had learned from it (p. vii).
>
> **Exploratory data analysis is detective work...** Confirmatory data analysis is judicial or quasi-judicial in character…. Unless the detective finds the clues, judge or jury has nothing to consider. **Unless exploratory data analysis uncovers indications, usually quantitative ones, there is likely to be nothing for confirmatory data analysis to consider.** (p. 1).

From the preceeding discussion of statistical analyses, it is apparent that there are many statistical methods and approaches to analyzing data. In order to establish the statistical methods to be used in analyzing abandoned mine discharge data for remining permitting and monitoring, it is necessary to consider the relationship between the characteristics of the sample data and the types of questions to be addressed in determining the baseline pollution load of the discharges. Sometimes, the characteristics of the available data do not lend themselves well to the type of statistical analysis which would be most appropriate to solve the problem. The type of statistical analysis which is: (1) appropriate to apply to a specific data set, and (2) desired or necessary to answer specific questions about the data depends upon numerous factors. These factors include:

- the sampling method,
- the number of observations included in the sample,
- the interval between observations in time (or space),
- the number of measurements performed (e.g., analyzing a water sample for 12 chemical constituents),

- the scale level of the data (i.e., nominal, ordinal, interval, ratio), and
- the frequency distribution of the data.

Univariate/Bivariate and Multivariate Analysis

Statistical analyses which evaluate a single variable are referred to as univariate analyses, while bivariate analyses evaluate the relationship between two variables. Multivariate statistical analyses concurrently evaluate the relationships among more than two variables. Statistical methods involving the frequency distribution of a variable(e.g., chi-square "goodness of fit" test, T-test of the significance of means, F-test of variance ratios) are examples of univariate statistical analyses. Linear regression and correlation (e.g., correlation coefficient (r), and coefficient of determination (r^2)) are examples of bivariate analyses, while multiple regression, factor analysis, principal components analysis, and cluster analysis are examples of multivariate analyses.

It is obvious that it will be very difficult, if not impossible, to use a univariate statistical method to solve a multivariate problem. For example, assume a mine drainage data set contains 100 water samples (i.e., number of observations, N = 100) which have been analyzed for 20 chemical constituents (i.e., number of parameters, p = 20), an N x p data matrix of 100 x 20 results, within which some of the parameters may be highly correlated or dependent upon each other (e.g., acidity, sulfate, and iron may vary in a closely associated pattern). If the problem to be solved is "how many independent sources of information are contained in the data matrix," a multivariate or "p-dimensional" problem exists that should be addressed with a multivariate statistical method such as principal components analysis or factor analysis. The evaluation of the shape of the frequency distribution of any or all of the 20 variates, in a univariate statistical context, may be an important part of the data analysis process, but it would not solve the multivariate problem.

As the level of sophistication and rigor of the statistical analysis increases from univariate through bivariate and multivariate to include some very powerful statistical methods such as time-series analysis, the requirements placed upon the quality of the data set increase in a corresponding manner. As described earlier, many parametric, univariate statistical methods are based upon the assumption that the sample data are normally distributed. Many bivariate statistical methods, such as linear regression which uses a least-squares method to determine a best fitting regression line, assume that the scatter of data points (when the two variates are plotted together) occurs in a uniform pattern, known as homoscedasticity. In general terms for correlation and regression analyses, this means that: (a) the scatter of the data points does not increase as the data values of the two variates increase, and (b) the data are normally distributed orthogonal to the regression line (i.e., within sections drawn perpendicular to the regression line at equal intervals along the line). Many multivariate statistical methods are based upon the assumption of joint normality of the data matrix (i.e., that all of the variates are normally distributed). Most multivariate statistical analyses are also greatly impeded by missing data (e.g., where 75 of 100 water samples were analyzed for 20 parameters, and the remaining 25 samples were analyzed for 12 parameters), as adjustments are made to the data matrix in order to enable the use of the matrix algebra necessary to mathematically solve the problem. The proper use of time series analysis generally requires a very large number of observations, equally spaced in time (i.e., equal intervals between observations), with no missing data.

Time Series Analysis

As stated earlier, the fundamental statistical problem to be addressed in determining baseline pollution load for remining permitting and monitoring purposes is how to summarize the natural variations in flow and water quality parameters before remining commences, in order to enable the separation of mining-induced changes in pollution load from natural seasonal variations in pollution load during and following remining operations. Conceptually, this is the type of statistical problem which is ideally solved by time-series analysis or a specialized area of time-series analysis, known as intervention analysis. However, the data quality requirements for these types of statistical analyses will exceed the available data for most remining cases, and to require remining permit applicants to collect sufficient data for these analyses would be an onerous and expensive task. The principles of time-series analysis will be briefly introduced here, and more fully explained in later Chapters.

The use of time series analysis in this report is chiefly for research purposes where adequate data exist. The results of research with time series analyses of relatively large mine drainage databases provide a better understanding of the behavior of abandoned mine discharges as they vary through time, and facilitate the application of a relatively simple quality control approach to the statistical analysis of the smaller sets of discharge data typically used in computing baseline pollution load in remining permits.

According to Vandaele (1983):

> A time series is a collection of observations generated sequentially through time. The special features of a time series are that the data are ordered with respect to time, and that successive observations are usually expected to be dependent. Indeed, it is this dependence from one time period to another which will be exploited in making reliable forecasts…. It also will be useful to distinguish between a time series process and a time series realization. The observed time series is an actual realization of an underlying time series process. By a realization we mean a sequence of observed data points and not just a single observation. The objective of time series analysis is to describe succinctly this theoretical process in the form of an observable model that has similar properties to those of the process itself. (p. 3)…. A time series model consisting of just one variable is appropriately called a univariate time series model. A univariate time series model will use only current and past data on one variable….. A time series model which makes explicit use of other variables to describe the behavior of the desired series is called a multiple time series model. The model expressing the dynamic relationship between these variables is called a transfer function model. The terms transfer function model and multiple time series model are used interchangeably. (p. 8)…. Finally, a special form of transfer function model is the intervention model. The special characteristic of such a model is not the number of variables in the model, but that one of the explanatory variables captures the effect of an intervention, a policy change, or a new law. (p. 9).

The use of intervention analysis to evaluate remining discharge data might be particularly appropriate providing that adequate data quality exists. One of the seminal works in intervention analysis is described in Tiao and Box and Hamming (1973) and Box and Tiao (1975) in which photochemical smog data from Los Angeles was analyzed in order to evaluate the effect of a new law requiring the reduction of reactive hydrocarbons upon the oxidant pollution level in the city. This is analogous to analyzing abandoned mine drainage pollution load data collected before, during and after remining in order to determine the effect of remining upon the level of the baseline pollution load in the presence of significant seasonal variations. An example of the use of intervention analysis of abandoned mine drainage data is the study by Duffield (1985) of the Arnot discharges, that are also featured in Chapter 4 of this report.

According to Box and Tiao (1975, p. 70):

> Data of potential value in the formulation of public and private policy frequently occur in the form of time series. Questions of the following kind often arise: "Given a known intervention, is there evidence that change in the series of the kind expected actually occurred, and, if so, what can be said of the nature and magnitude of the change?" … In the examples quoted, however, the data are in the form of time series, in which successive observations are usually serially dependent and often nonstationary and there may be strong seasonal effects. Thus, the ordinary parametric or nonparametric statistical procedures which rely on independence or special symmetry in the distribution function, are not available nor are the blessings endowed by randomization.

Intervention analysis and other methods of time series analysis are very powerful statistical tools which would be desirable and useful in evaluating baseline pollution load data, but these types of statistical analyses will usually be inappropriate for remining permitting due to inadequate data availability and data quality. Therefore, it was necessary to develop a data analysis algorithm which recognized or allowed for the use of time-series analyses, but did not require the routine use of these statistical methods in order to answer the desired questions about the remining discharge data.

A flow chart outlining the data analysis algorithm for determining the baseline pollution load is shown in Figure 3.1. The algorithm includes evaluations of data quality, univariate statistical analyses, bivariate statistical analyses and time series analyses methods to establish quality control limits. The algorithm includes steps to evaluate the normality of the frequency distribution and transform the data if the distribution is not normal (i.e., positively skewed); however, the use of the statistical methods in the algorithm does not require the distribution to be normal. The algorithm contains elements of parametric statistical analysis, but it is primarily based upon order statistics and non-parametric statistics.